Transferability of Travel Survey Data: A Household Travel Data Simulation Tool

Abolfazl (Kouros) Mohammadian
Associate Professor
Department of Civil and Materials Engineering
University of Illinois at Chicago

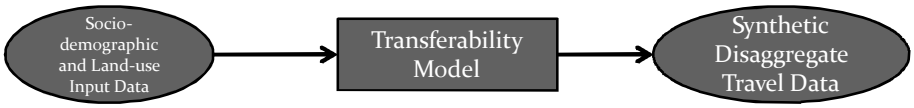
UIC University of Illinois at Chicago

TMIP Webinar
January 25, 2010

Objectives

- Examine the feasibility of data transferability approach as an alternative to costly travel surveys
- Develop a comprehensive travel data transferability model
- Develop a software tool to facilitate travel data transferability
  - simulate synthetic household level disaggregate travel data
- Test and validate the model
  - Two case studies: Des Moines, IA and New York State
- Examine the model for sensitivity analysis
  - scenarios such as changes to the demographics, aging population, and investments in the education system.

UIC Transferability Tool



Input Data:

- Socio-Demographic Information of Households
  - Usually comes from population synthesizer
- Land-use Characteristics of Residential locations
  - Embedded in the software for all census tracts in US

Input Data 1:

- Socio-Demographic Information of Households
  - Education Level
  - Type of Job
  - Type of Ethnicity
  - Household Size
  - Household Income
  - Number of Vehicles
  - Number of Workers
  - Number of Adults

Input Data 2:

- Land-use Characteristics:
  - Housing density, Employment density, Population density
  - Transit usage (Compiled from CTPP 2000)
  - Travel-Time Index (from TTI report)
  - Intersection density
  - Road density
  - Pedestrian friendly environment (Block size)

UIC Transferability Model

- NHTS 2001 households were clustered into 11 clusters, according to their life-style
- At each cluster, many distributions were fitted on travel attributes, and the parameters of best ones are used to generate travel attribute for other regions
- A neural network model is used to replicate the clustering
- Monte-Carlo simulation is then used to generate travel attributes of the households (Drawing from the selected distribution)
- Updating methods are suggested in case more information of the application region is available (local sample or expert's opinion)
- For more information please refer to:

Mohammadian, A. and Y. Zhang. Investigating the Transferability of National Household Travel Survey Data, in Transportation Research Record: Journal of the Transportation Research Board, No. 1993, TRB, 2007, pp.67-79.

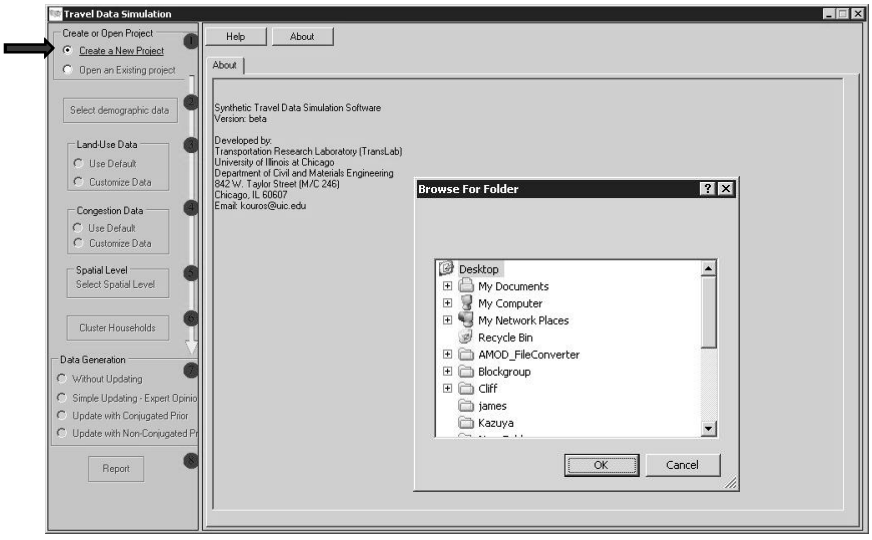
Output:

- Nine travel attributes for every household
  - Total number of trips
  - Number of auto trips
  - Number of mandatory trips
  - Number of maintenance trips
  - Number of discretionary trips
  - Number of tours
  - Average number of trips per tours
  - Average trip miles traveled
  - Average commute distance
- Can be aggregated at any spatial level

Using Transferability Software: 8 Steps

- Create/open a project
- Import socio-demographic information
- Use provided land-use data or import your own
- Use provided congestion index or import your own data
- Select spatial level of each type of input data
- Cluster the households
- Generate Travel attributes
- Generate reports

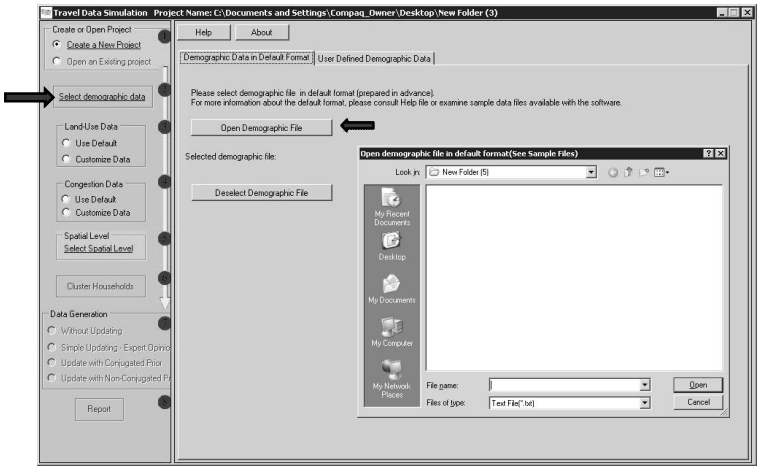
Step 1: Create/open a project



Step 2: Import socio-demographic information

- Includes a list of households, their residential location and socio-demographic information
- Usually compiled from population synthesizers' output
- Text file format

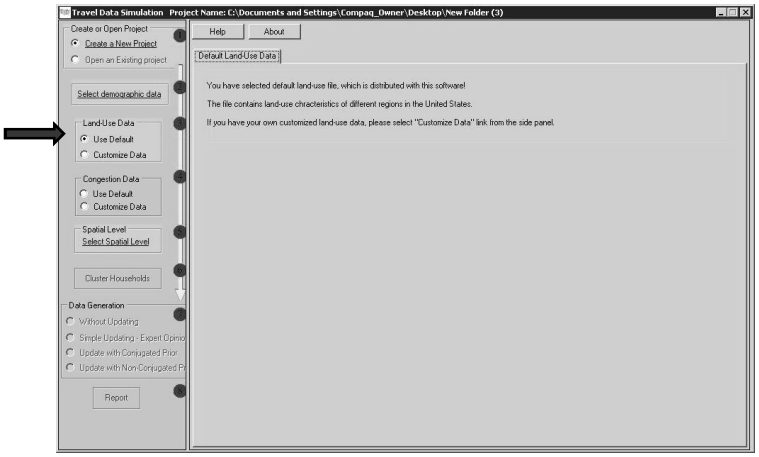
Step 2: Import socio-demographic information



Step 3: Use the land-use data provided

- Currently land-use characteristics at Census Tract (CT) level are embedded in the software.
- For most applications, the embedded information can be used.

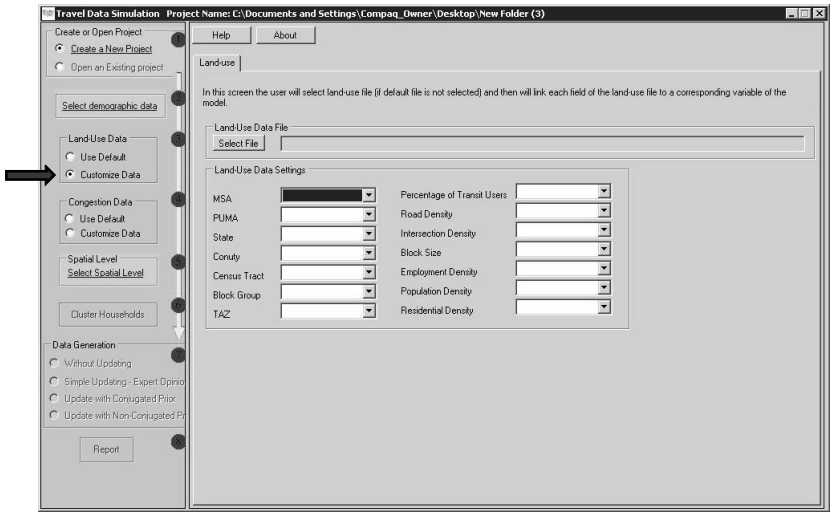
Step 3: Use the land-use data provided



Step 3: Import your own land-use data

- If more recent land-use information is available, it can be imported and used in the analyses.
- Also if further detailed information (e.g. Block Group level characteristics) is available, it be imported as well.

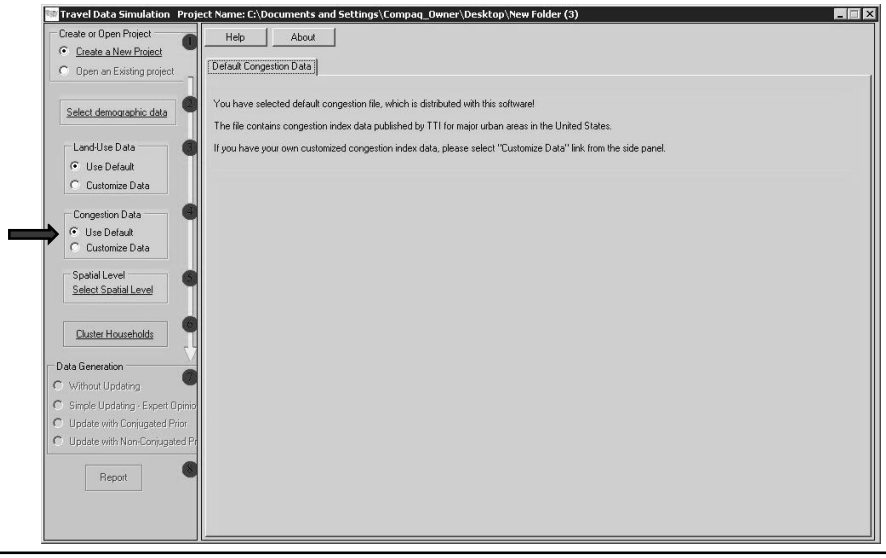
Step 3: Import your own land-use data



Step 4: Use provided congestion index

- For most applications, there is no need to import any data.
- Embedded information (from TTI annual mobility report) can be used.

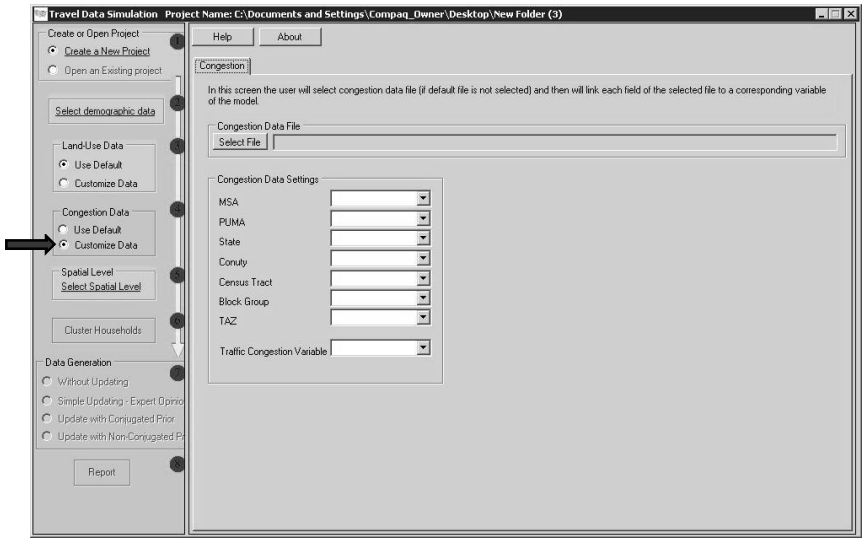
Step 4: Use provided congestion index



Step 4: Import your own congestion index

- The provided file, includes Travel Time Index of 85 urban areas (MSA). If more recent data is available, it can be imported.
- Also if further detailed information (e.g., Census Tract level) is available, it can be imported as well.

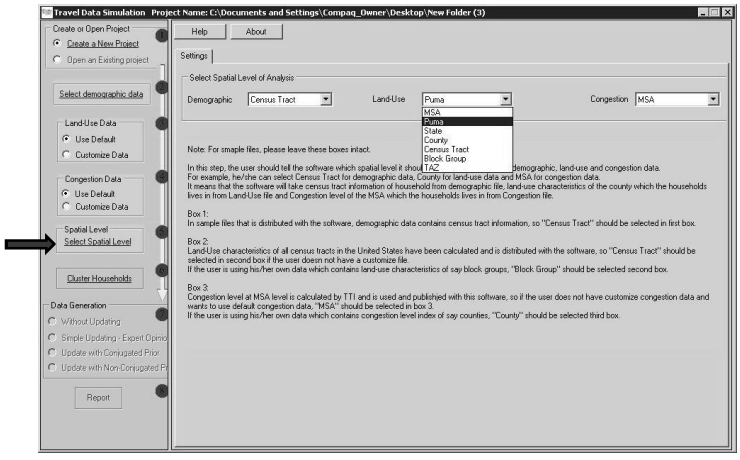
Step 4: Import your own congestion index



Step 5: Specifying spatial level of input data

- The default spatial level of data are:
  - Demographic data: Census Tract
  - Land-Use data: Census Tract
  - Congestion: Urban Area
- If other spatial levels are used, they should be identified in this step

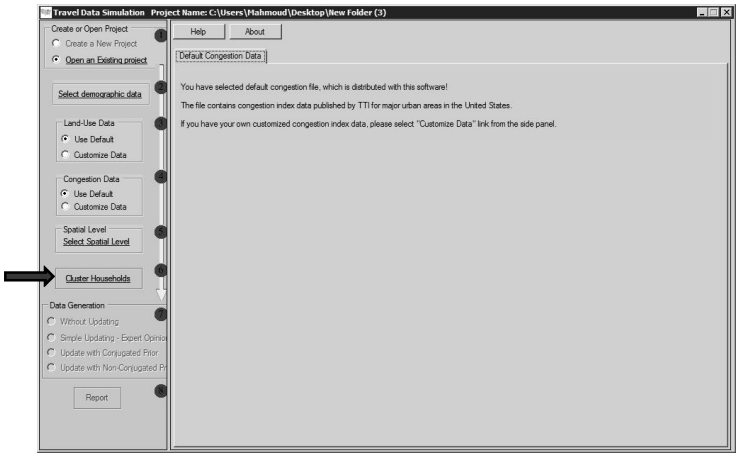
Step 5: Specifying spatial level of input data



Step 6: Clustering the Households

- In this step, the software combines the input data to assign cluster membership (lifestyle) to each household.
- The output is a text file, including a list of households along with their characteristics and cluster numbers.

Step 6: Clustering the Households



Step 7: Generating Disaggregate Travel Data

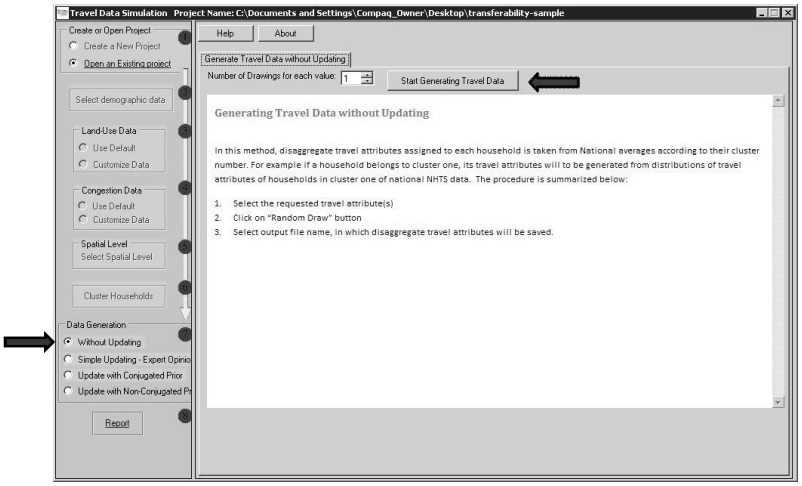
- Predefined distributions are used to generate synthetic travel data
- Four options for generating travel data are offered in the software:
  1. Without updating
  2. Updating parameters of distributions using an expert's opinion
  3. Updating parameters of distributions using mean and sample size of a local sample and Bayesian updating method (Conjugated)
  4. Updating parameters of distributions using a local sample and Bayesian updating method (Non-Conjugated)

Step 7: Generating Disaggregate Travel Data

- Method 1: Using predefined distributions (without updating)
  - For each travel attribute and across the 11 clusters, the program uses a random number generator to draw a set of random numbers from appropriate distributions and assigns them to the households.
  - Best fitted distributions are embedded into the software.

Step 7: Generating Disaggregate Travel Data

- Method 1: Using predefined distributions (without updating)

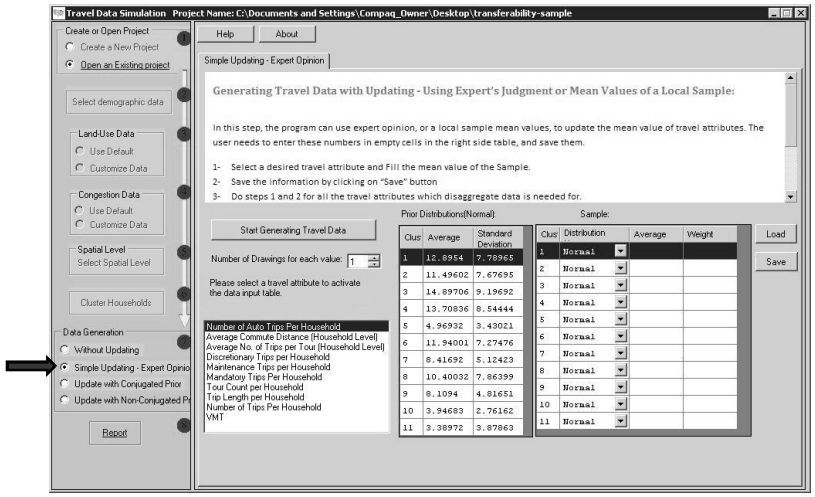


Step 7: Generating Disaggregate Travel Data

- Method 2: Expert's opinion updating
  - Just an estimate of average values of travel attributes, at each cluster, is needed for the application region
  - They are used to update the parameters of the distributions that were fitted on national level travel data.
  - Travel attributes are drawn from normal or gamma distribution (Gamma fits reasonably well on reported travel attributes)

Step 7: Generating Disaggregate Travel Data

- Method 2: Expert's opinion updating

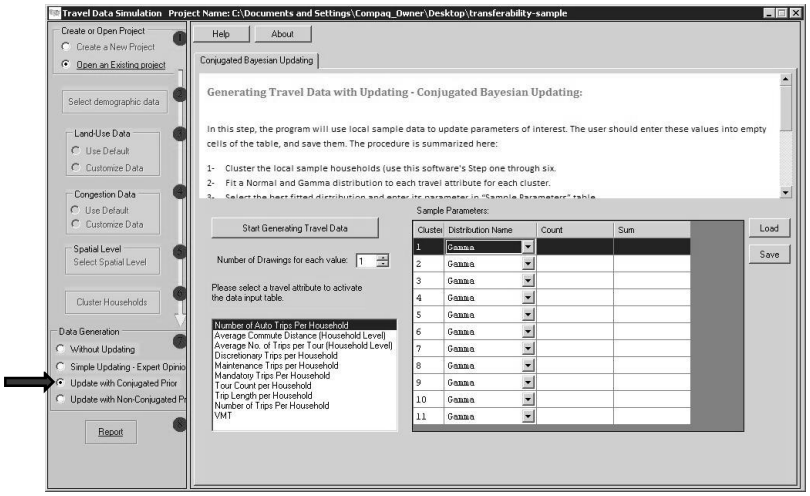


Step 7: Generating Disaggregate Travel Data

- Method 3: Bayesian updating -Conjugated
  - Bayes theorem is used to take advantage of a local sample to update the parameters of the distributions that were fitted on NHTS 2001 national travel data.
  - Bayesian updating for gamma and normal distributions have closed forms and could be incorporated into the software.
  - Just basic statistics (average and sample size) of the local sample is needed to apply this method.

Step 7: Generating Disaggregate Travel Data

- Method 3: Bayesian updating –conjugated

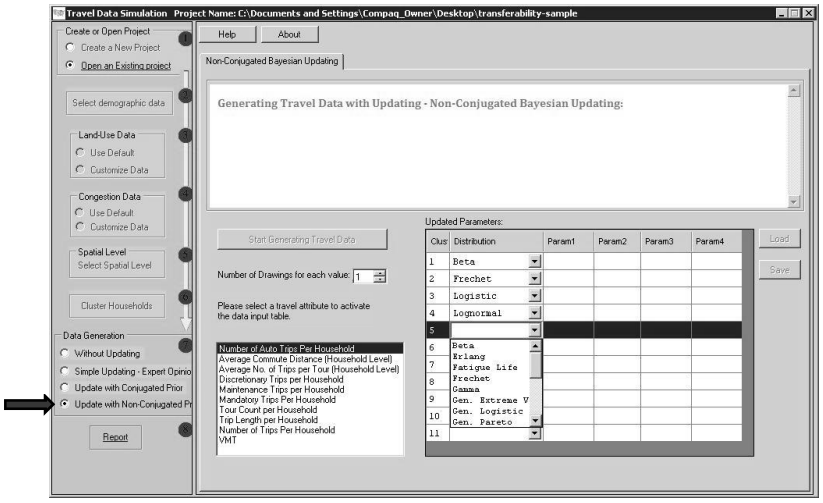


Step 7: Generating Disaggregate Travel Data

- Method 4: Bayesian updating –Non Conjugated
  - For most distributions there is no closed form formula for updating the parameters, therefore advanced Bayesian updating software tools should be employed.
  - The user needs to update the parameters of the distributions using a professional software (e.g., WinBUGS), and enter the results into the transferability software in order to generate disaggregate travel data.

Step 7: Generating Disaggregate Travel Data

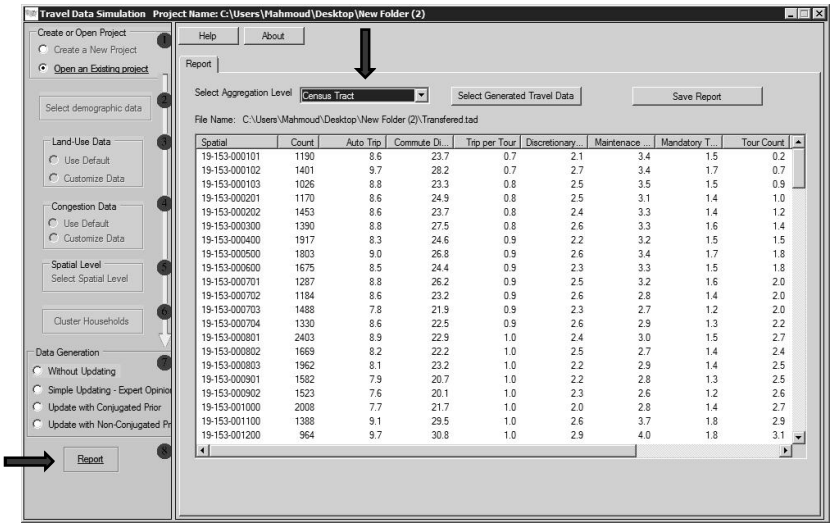
- Method 4: Bayesian updating –non conjugated



Step 8: Generating Reports

- The output of previous step is a text file including a list of households along with:
  - Their socio-demographic attributes
  - Land-use characteristics of their residential locations
  - Their lifestyle (i.e., cluster number)
  - Their synthetic travel data
- The software can generate aggregate reports at any spatial level (state, county, census tract, block group, and TAZ)

Step 8: Generating Reports



Validation

- The model was used to generate travel attributes of two regions (a mid-size and a large area):
  - Des Moines, IA and the state of New York
- The results were validated against corresponding NHTS 2001 add-ons

Des Moines Observed, Transferred, and Updated Travel Attributes (first 3 clusters)

Cluster	Case	Number of Households	Auto Trip	Commute Distance	Trips per Tour	Discretionary Trips	Maintenance Trips	Mandatory Trips	Tour Count	Trip Length	Trip Rate
1	Observed	68	12.4	23.1	1.3	4.2	2.8	1.9	11.0	141.3	13.8
	Transferred	10278	7%	9%	9%	-13%	**48%**	13%	-1%	**-20%**	10%
	Updated	10278	1%	-5%	1%	-5%	20%	2%	0%	-16%	1%
2	Observed	138	10.2	18.9	1.3	2.9	2.8	1.5	8.5	94.3	11.2
	Transferred	39153	13%	**78%**	4%	5%	**30%**	**32%**	11%	**22%**	14%
	Updated	39153	-1%	12%	1%	-2%	1%	9%	-2%	3%	-1%
3	Observed	104	15.6	27.4	1.4	4.9	4.2	2.3	12.7	158.4	17.3
	Transferred	11214	-4%	13%	1%	-14%	14%	5%	-3%	-14%	-1%
	Updated	11214	-4%	-7%	-1%	-7%	2%	-5%	-1%	-9%	-2%

New York Observed, Transferred, and Updated Travel Attributes (first 3 clusters)

Cluster		Number of Households	Auto Trip	Commute Distance	Trips per Tour	Discretionary Trips	Maintenance Trips	Mandatory Trips	Tour Count	Trip Length	Trip Rate
1	Observed	738	12.3	34.6	1.4	3.6	3.9	2.1	10.5	121.1	14.5
	Transferred	190,171	7%	-27%	-1%	3%	7%	2%	4%	-7%	3%
	Updated	190,171	3%	-1%	-1%	4%	4%	2%	4%	0%	3%
2	Observed	889	10.0	24.2	1.4	2.6	3.4	1.8	8.5	94.5	11.8
	Transferred	444,173	14%	**38%**	-1%	16%	9%	8%	10%	20%	9%
	Updated	444,173	-1%	2%	1%	-2%	-2%	3%	-2%	-2%	-2%
3	Observed	801	13.1	31.0	1.4	3.6	4.2	2.3	11.4	129.5	15.7
	Transferred	255,182	14%	0%	-1%	15%	13%	5%	8%	7%	9%
	Updated	255,182	0%	-2%	-1%	2%	-1%	-1%	0%	-2%	-1%

Software

- The software is available for download from:
  - http://www.nationaltraveldata.com
- Two synthetic population sample files are also available for download to test the software

Population Synthesizer

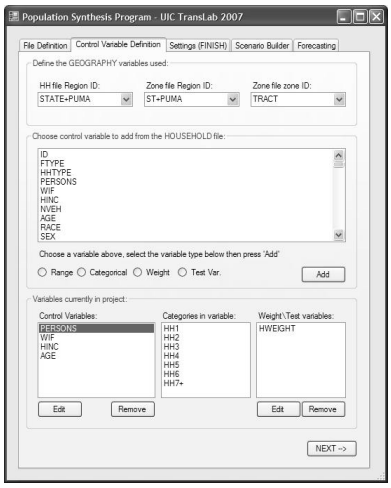
Synthesis Methodology

- Procedure (for each zone, i.e. block group, census tract, etc.:
  - Get joint distributions from sub-region level (PUMS) data
  - Fill zonal joint distributions (household and person) from sub-region
  - Fit joint distribution to zonal marginals (SF_3) for hh and person using IPF procedure
  - For each household in the PUMS sample (drawn randomly):
    - attempt to add x times with probability p
    - x = number of remaining households to create with the same type as the current household (from IPF)
    - P = household selection probability weighted by fit to person distribution (shown below)
    - Update x and p each time household (and persons in household) are added
  - Ensures fit to both the household and person distributions with minimal iteration through household list (only one iteration required if rounding/integerization effects ignored)

Population Synthesizer

Base Population Synthesis Program

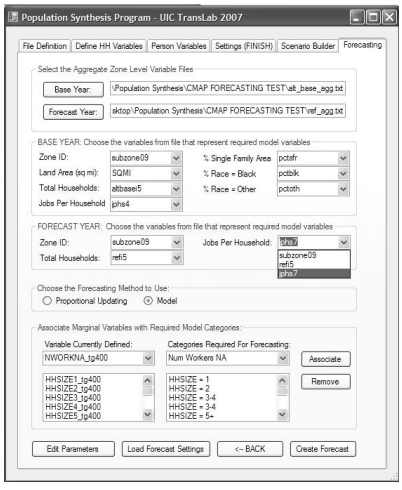
- Allows any geography and control variable for which data exists
- Up to nine controls at household and person level simultaneously
- Synthesis routine:
  - IPF at household/person level
  - Household selection weighted based on person fit
- Output statistics include:
  - Marginal fit for controlled / uncontrolled variables
  - Statistical tests on synthesized joint distributions
- Fast visualization of results



Population Synthesizer

Forecasting Control Variables

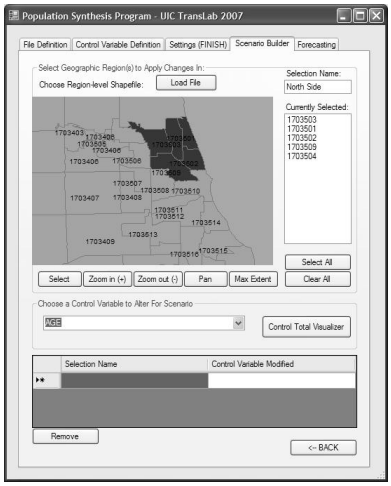
- Input base and forecast year required zonal data
- Link control variable categories to forecast categories
  - 4 HHsize, 3 numworkers
- Generate forecast marginals:
  - Proportional updating, or
  - Forecast model

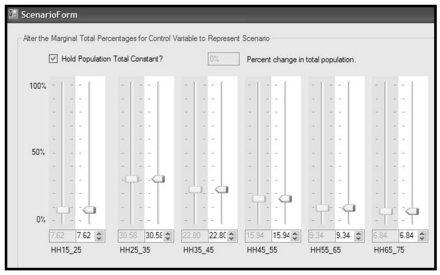


Population Synthesizer

Scenario Definition

- Select sub-regions to apply changes
- Select control variable to modify
- Adjust variable marginal distribution
- Multiple selections, modified variables allowed





Population Synthesizer

Benefits of Population Synthesis Program

- Very general:
  - Not restricted in terms of geographic applications
  - Any control variables can be used (given appropriate data)
  - Other applications besides Household/Person synthesis
- Multilevel controls:
  - More control over synthesized population, less person-level error
- Implementation of efficient selection procedure:
  - Reduces run-time to <2 sec/zone (in many instances)
    - for average of 2,000 synthesized households per zone
- Standard output statistics allow for comparisons of different synthetic populations, fine-tuning of final result

Questions?